(12) UK Patent Application (19) GB (11) 2 226 718 (13) A

(43) Date of A publication 04.07.1990

(21) Application No 8925698.6

(22) Date of filing 14.11.1989

(30) Priority data
(31) 8826927 (32) 17.11.1988 (33) GB

(51) INT CL[5]
G11B 27/00, H03L 7/00

(52) UK CL (Edition K)
H3A AA
G5R RB788 RB81

(56) Documents cited
GB 2020080 A

(58) Field of search
UK CL (Edition J) G5R RB81, H3A AA ARX ASX AXX, H4P PDCSS
INT CL[4] G11B, H03B, H03L

(71) Applicant
British Broadcasting Corporation

(Incorporated in the United Kingdom)

Broadcasting House, London, W1A 1AA, United Kingdom

(72) Inventors
David Graham Kirby
Andrew James Mason

(74) Agent and/or Address for Service
Reddie & Grose
16 Theobalds Road, London, WC1X 8PL, United Kingdom

(54) Aligning two audio signals

(57) In a method for aligning two audio signals A and B, e.g. for automatic editing between recordings, repeated measurements are made of the similarity between the two signals and an optimum time offset for aligning the signals is determined. Sample sections of the two signals around the 'out' and 'in' points chosen by the user are outputted by facility 11 and sub-sections are analysed by Fast Fourier Transform circuits 12, 13 to derive a corresponding series of frequency spectra. Peaks in the correlation function performed at 14 between the two spectra are detected at 15 and from the position of the peaks, the best shift to apply to one of signals A, B to bring them into time alignment is deduced at 16. The hardware 12-16 may be replaced by a computer or microprocessor.

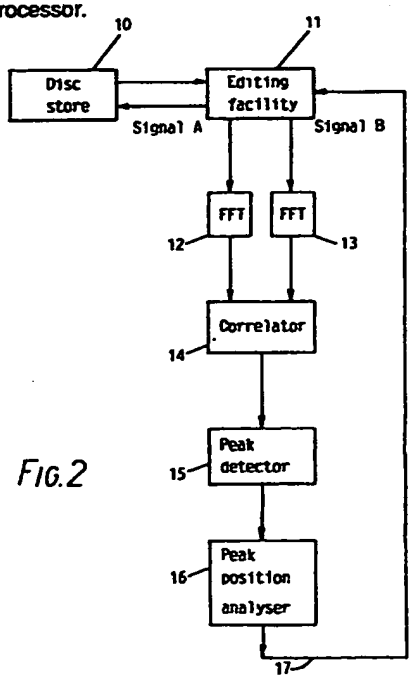


FIG.2

GB 2 226 718 A

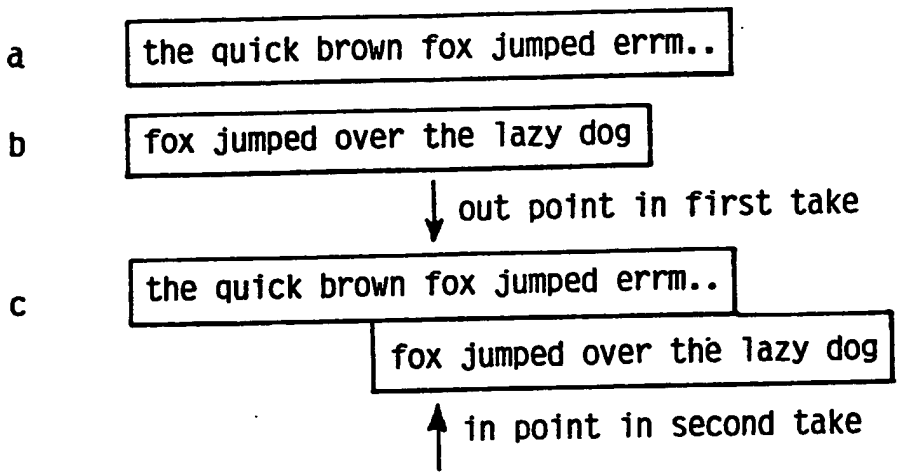

a
the quick brown fox jumped errm..
b
fox jumped over the lazy dog
out point in first take
c
the quick brown fox jumped errm..
fox jumped over the lazy dog
in point in second take

FIG. 1

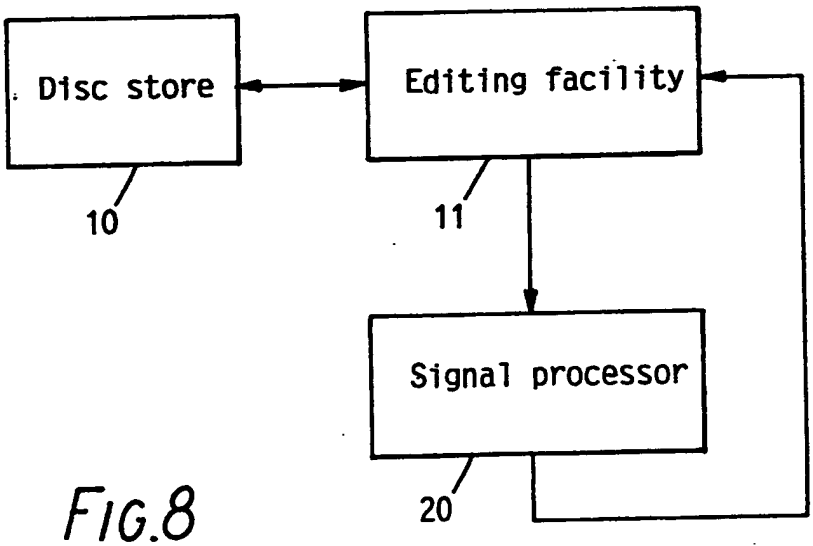

Disc store
Editing facility
10
11
Signal processor
20

FIG. 8

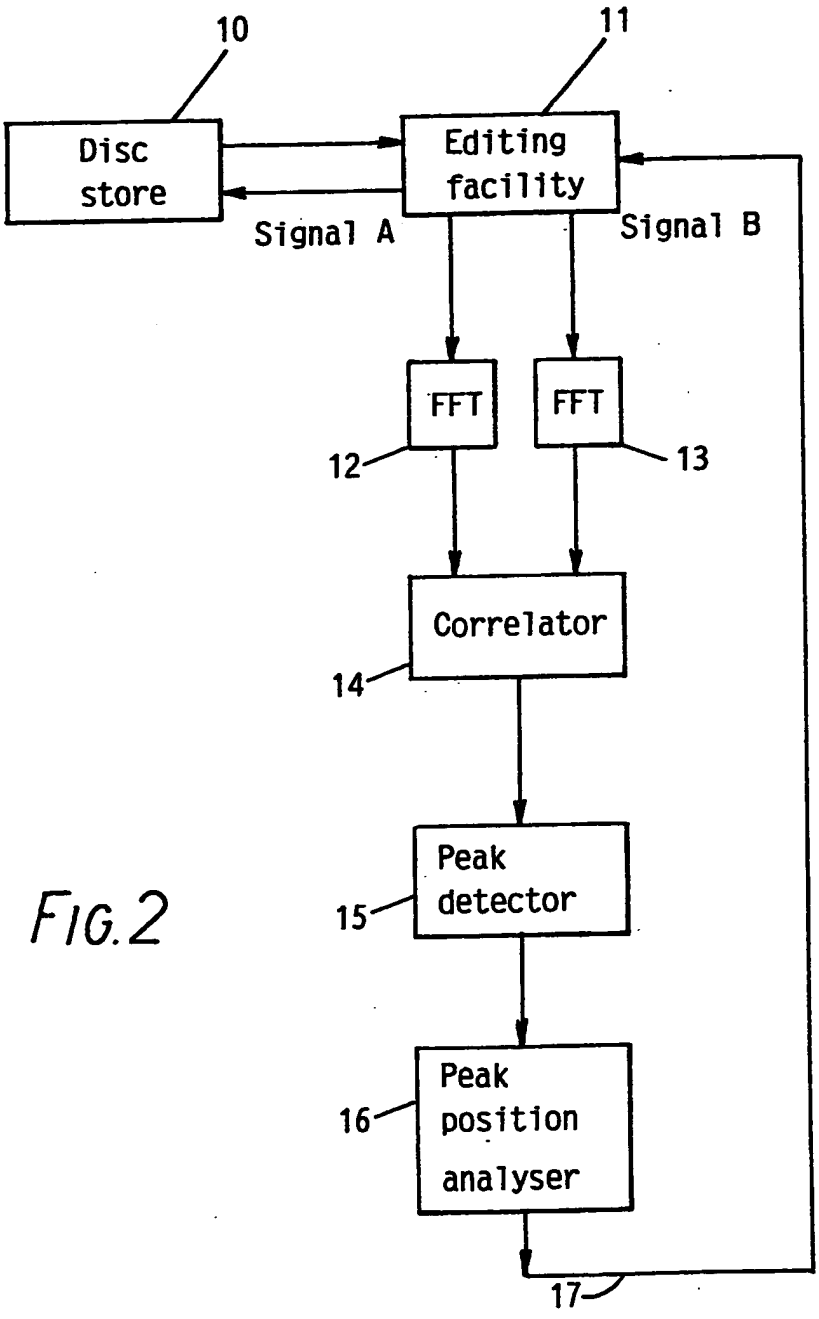

10
11
Disc store
Editing facility
Signal A
Signal B
FFT
FFT
12
13
Correlator
14
Peak detector
15
Peak position analyser
16
17

FIG. 2

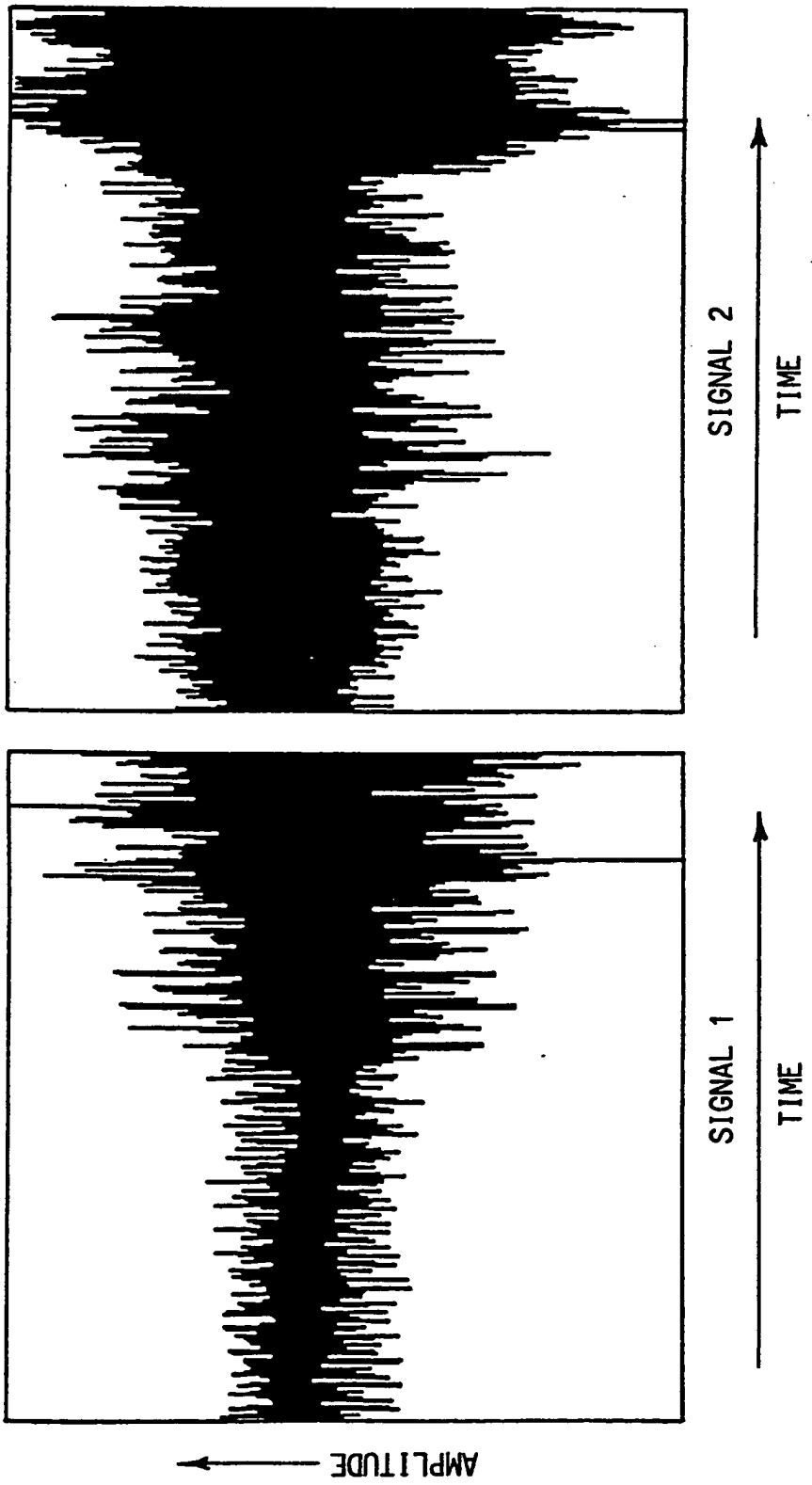


FIG.3

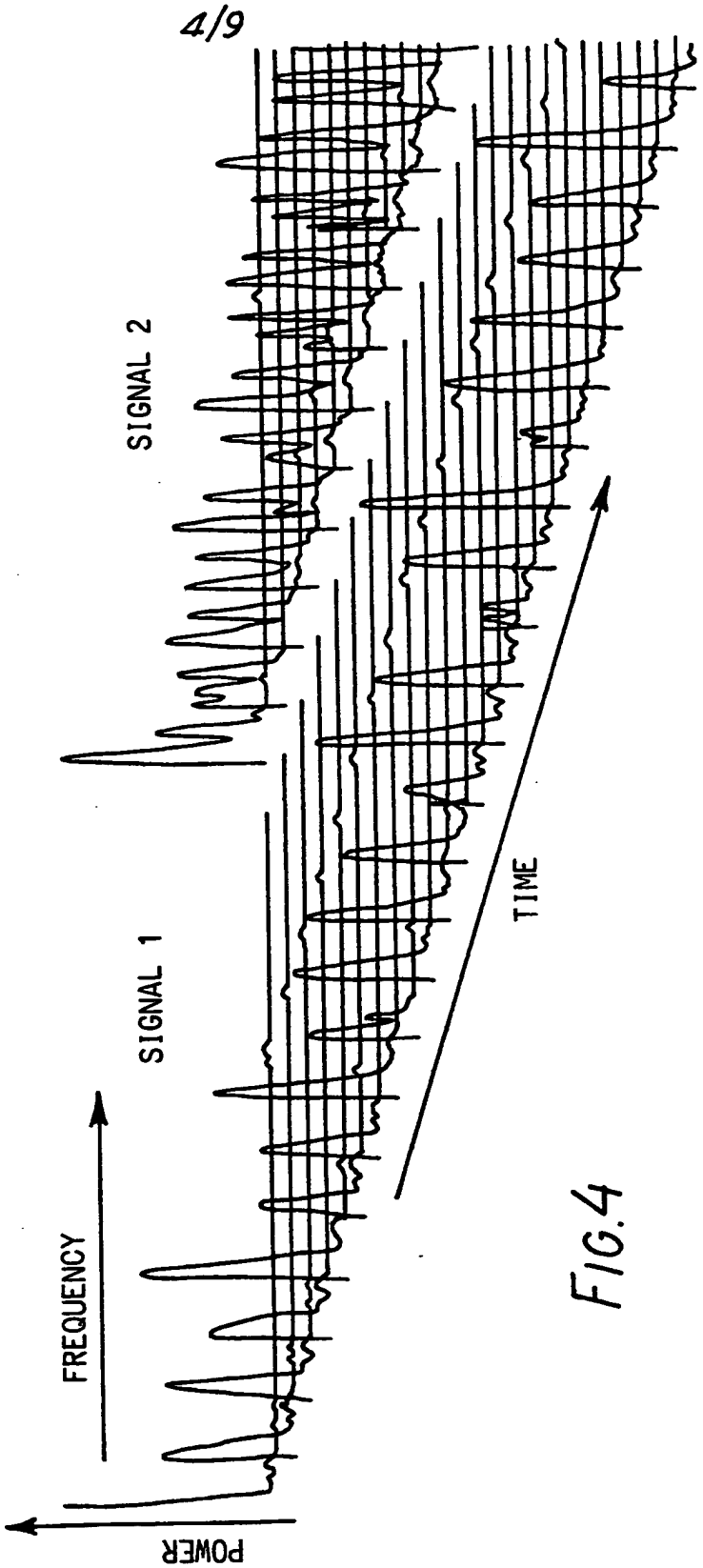
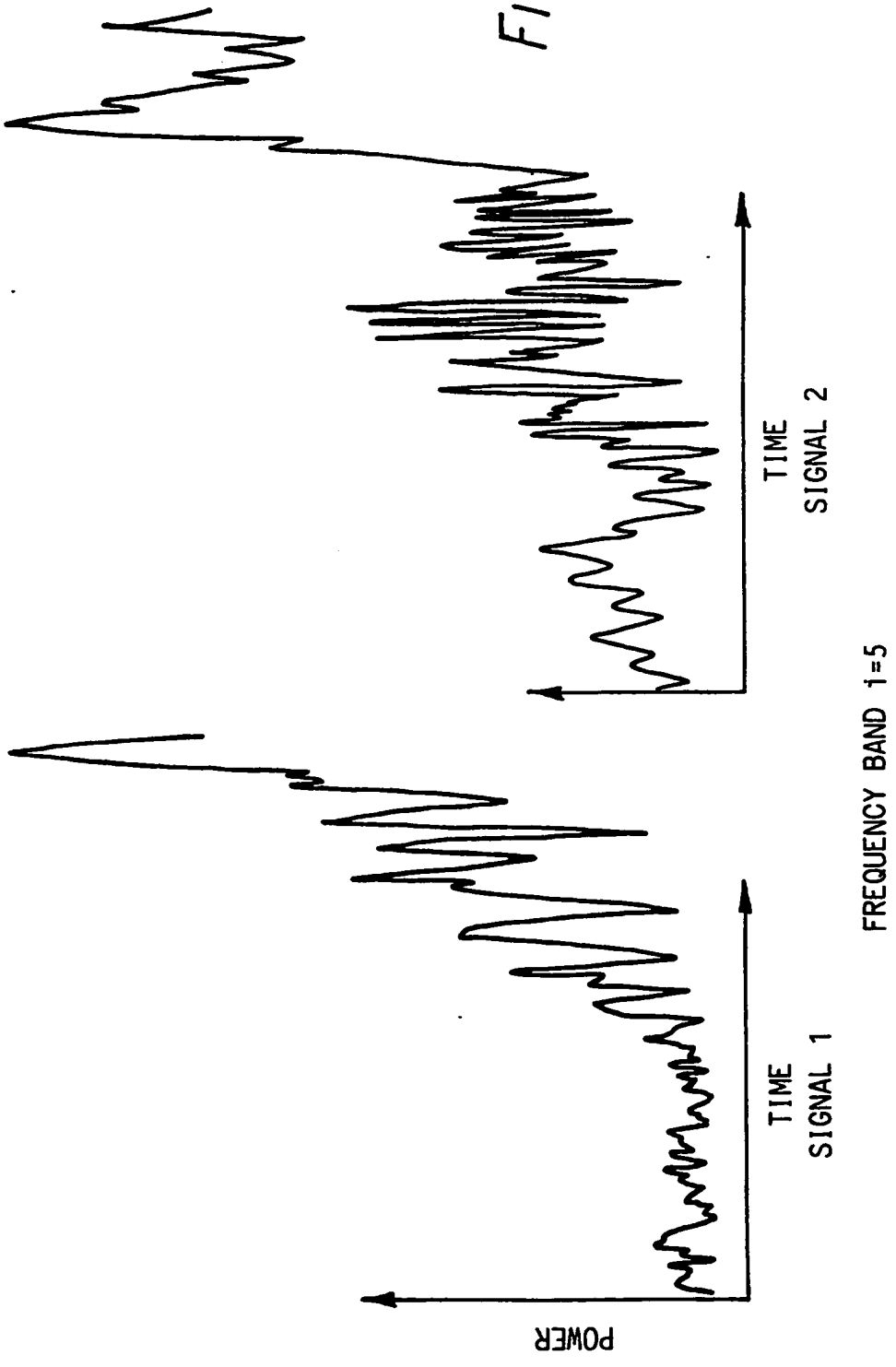


FIG.4

FIG.5



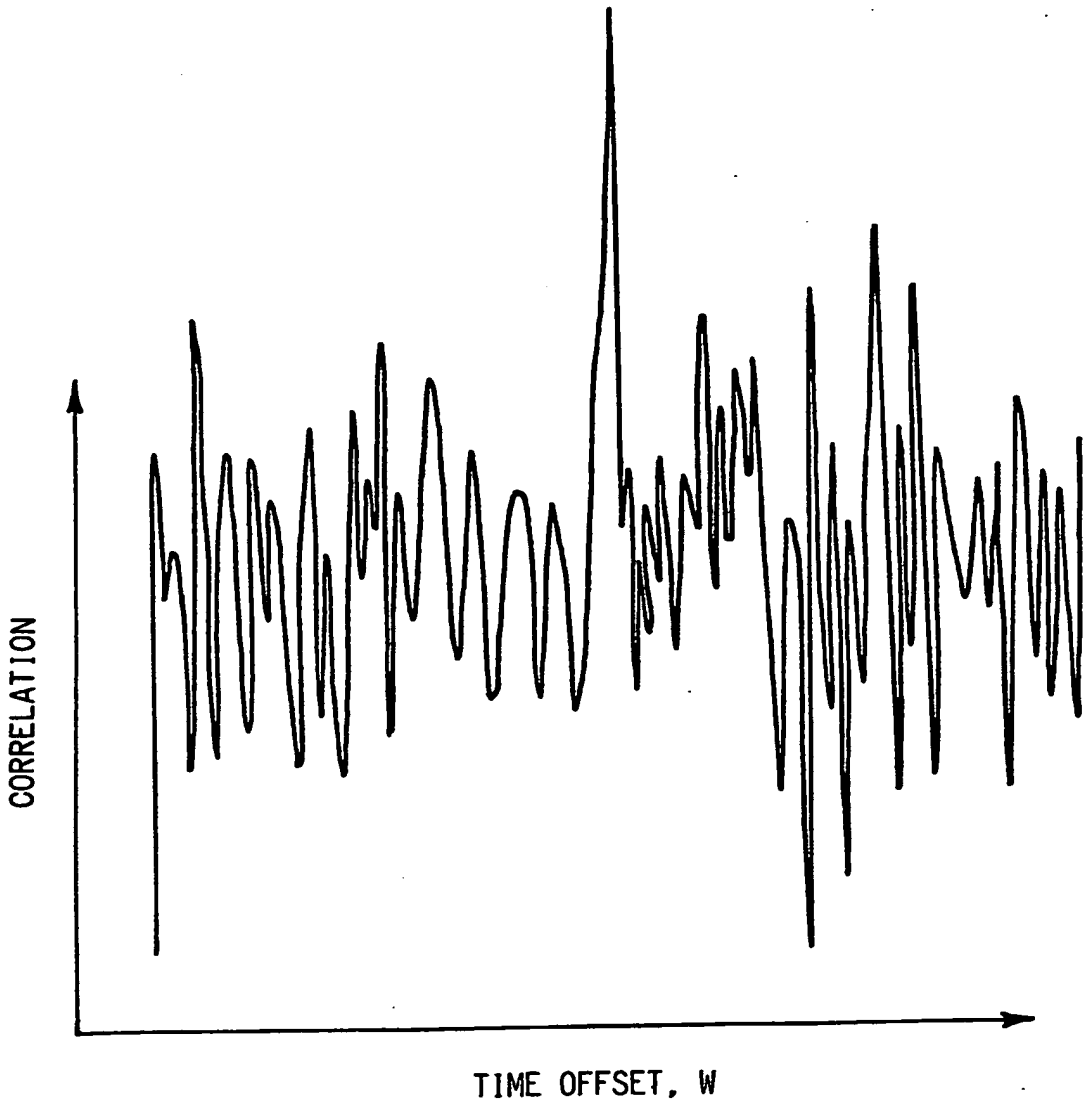


FIG.6

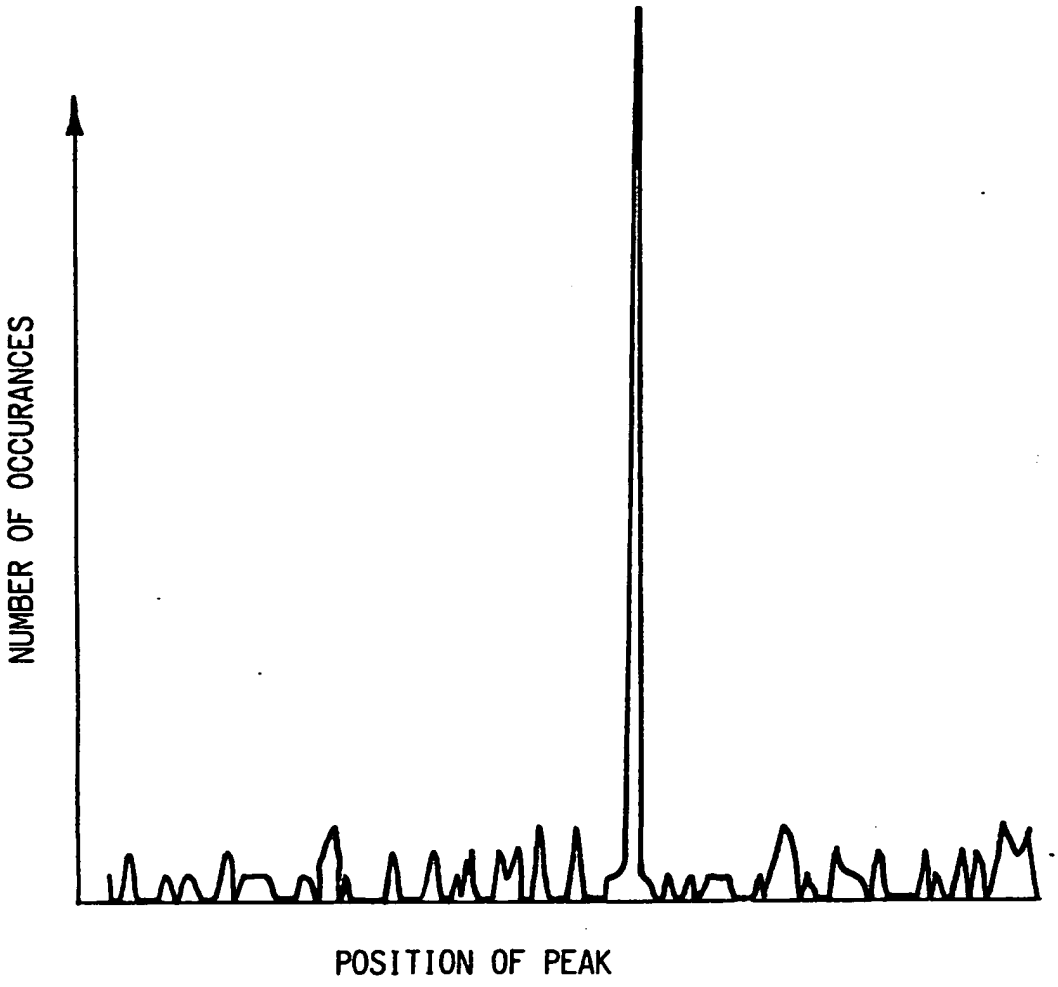


FIG. 7

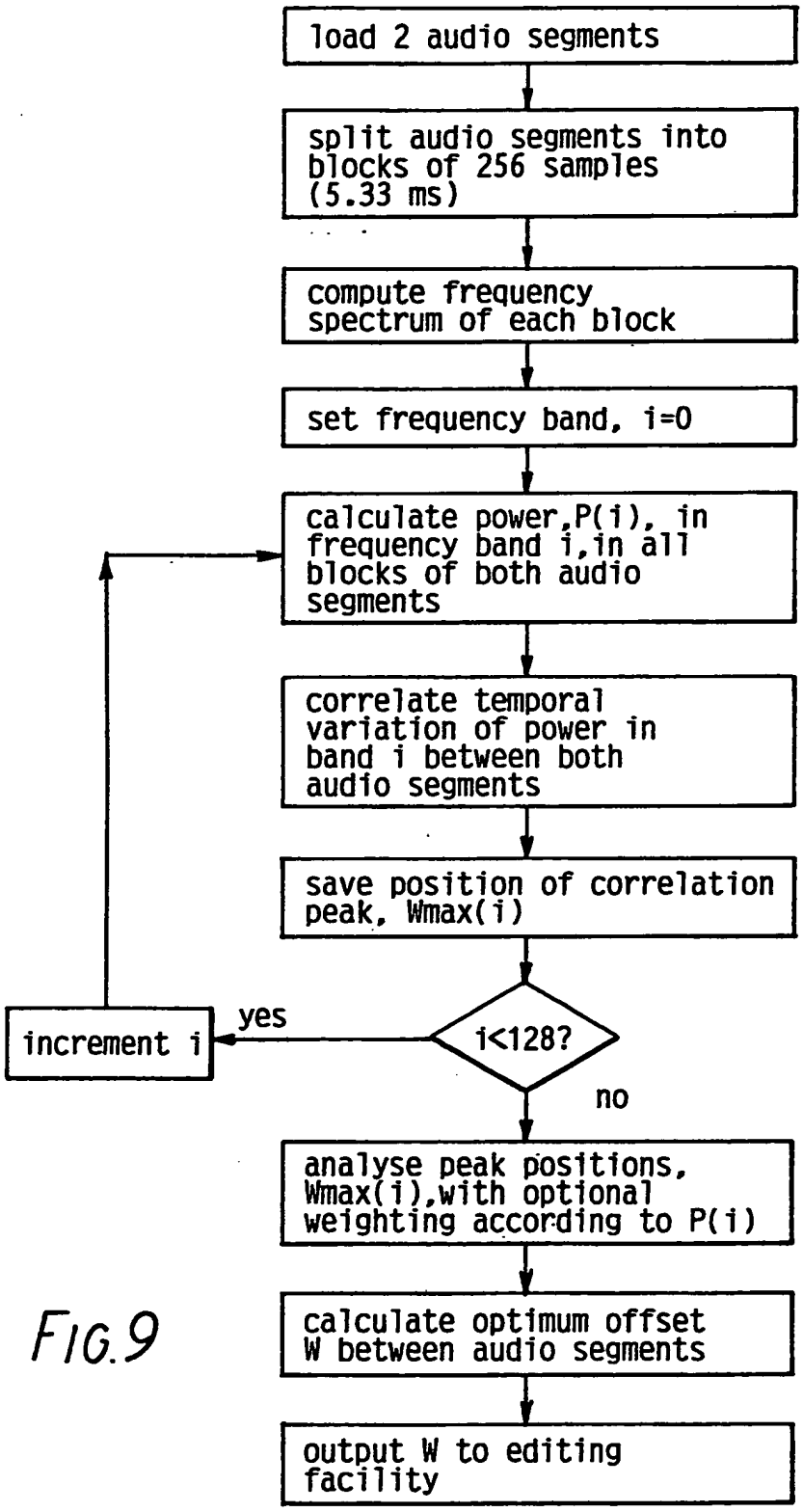

load 2 audio segments
split audio segments into blocks of 256 samples (5.33 ms)
compute frequency spectrum of each block
set frequency band, i=0
calculate power,P(i), in frequency band i,in all blocks of both audio segments
correlate temporal variation of power in band i between both audio segments
save position of correlation peak, Wmax(i)
i<128?
yes
increment i
no
analyse peak positions, Wmax(i),with optional weighting according to P(i)
calculate optimum offset W between audio segments
output W to editing facility

FIG.9

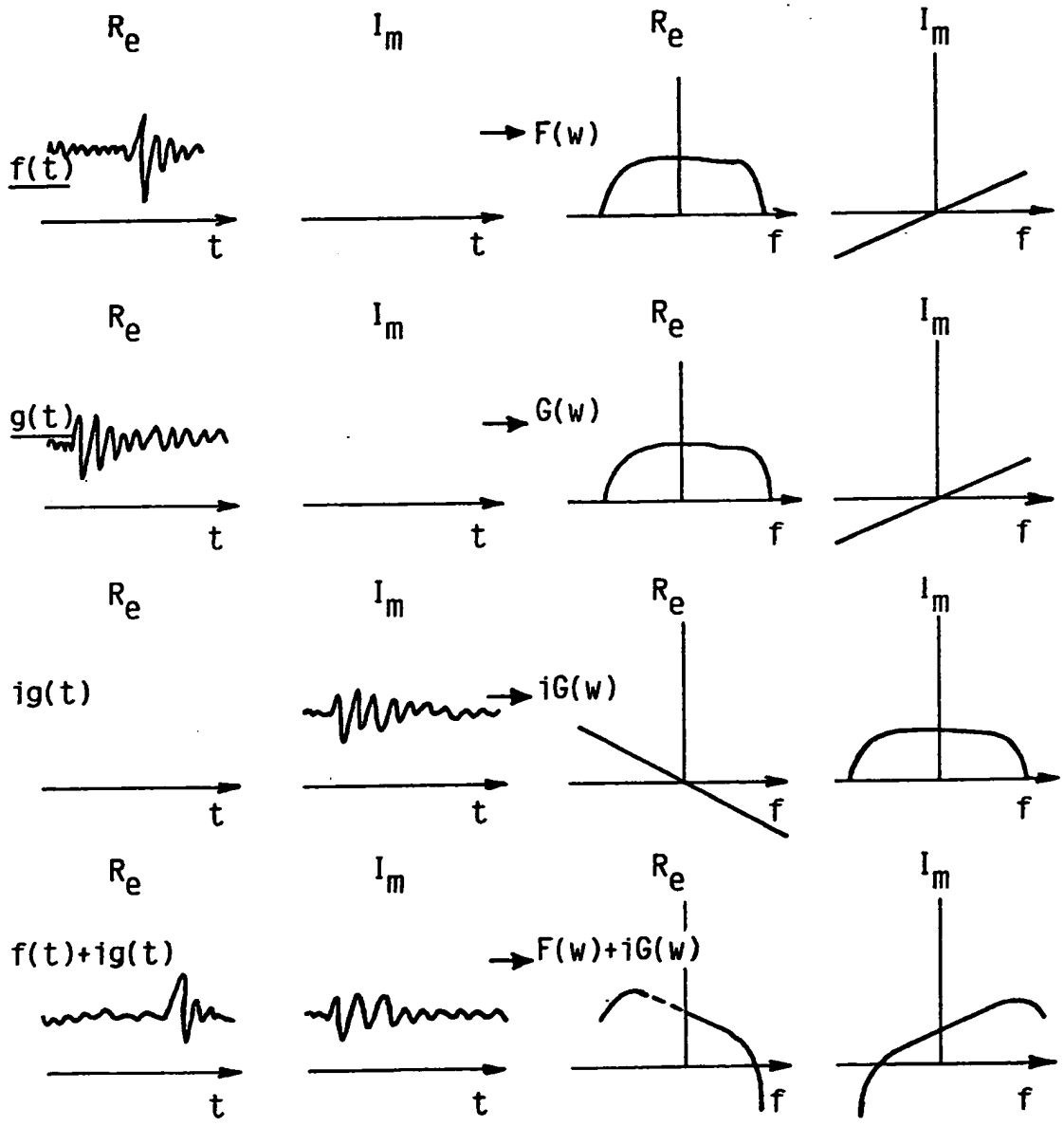

Re
Im
Re
Im
f(t)
F(w)
t
t
f
f
Re
Im
Re
Im
g(t)
G(w)
t
t
f
f
Re
Im
Re
Im
ig(t)
iG(w)
t
t
f
f
Re
Im
Re
Im
f(t)+ig(t)
F(w)+iG(w)
t
t
f
f

FIG.10

ALIGNING TWO AUDIO SIGNALS IN TIME, FOR EDITING

This invention relates to the field of audio recording and specifically to a method for aligning two audio signals in time, for instance for automating the adjustment of edits between recordings to obtain a high quality edit without manual intervention.

BACKGROUND OF THE INVENTION

In audio recording work it is frequently necessary to edit material together to remove mistakes or intrusive noises, for example. This is traditionally carried out by locating a suitable point in a first audio recording prior to the error and then finding the matching point in a second recording of the same material. The edit is then carried out between these two points, joining the former to the latter to remove the flawed section of the material. The perceived quality of the resulting edited audio material is critically dependent on the accuracy with which these two edit points are located.

The timing of the second recording relative to the first at the instant of the edit will determine whether audio material is repeated or lost as the edit is replayed. This will affect the extent to which the edit is imperceptible when replayed. Traditionally the location of these edit points is carried out by listening to the audio recordings at low speed and identifying the appropriate instants so as to align the two recordings in time. This is a skilled operation for which considerable experience is required.

SUMMARY OF THE INVENTION

The object of the present invention is to provide a method of aligning two audio recordings in time such as for the purpose of performing an edit between them thereby eliminating the need for the manual adjustment of the timing of one recording relative to the other.

The invention is defined in the appended claims to which reference should now be made.

Briefly described in its preferred embodiment, the invention uses a method of comparing the similarity of two audio signals in a multiplicity of frequency bands with varying time offsets between the two signals. The similarity measurements are then used to derive a measurement of the relative timing of the two audio signals and hence the time offset which must be applied to one of the audio signals to bring it into time alignment with the other.

BRIEF DESCRIPTION OF THE DRAWINGS

In order that the manner in which the foregoing can be understood in detail, a particularly advantageous embodiment thereof will be described with reference to the accompanying drawings, in which:-

Figure 1 is a representation of the editing process indicating the necessary time alignment of the two audio recordings and the position of the edit between them;

Figure 2 is a block circuit diagram of apparatus for aligning two audio recordings in time embodying the invention;

Figure 3 is a representation of two audio signals varying with time;

Figure 4 is a representation of the frequency spectra of the two signals varying in time;

Figure 5 is a representation of the power contained in a frequency band of each of the two signals varying with time;

Figure 6 is a representation of the correlation function of the two functions represented in Figure 5;

Figure 7 is a histogram of the position of the peaks in the correlation functions (one of which is shown in Figure 6) of all the frequency bands of the signals;

Figure 8 is a block hardware diagram of a computer-based embodiment of the invention;

Figure 9 is a flowchart illustrating the processor operations in the system of Figure 8; and

Figure 10 illustrates the fast Fourier transform operation employed.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Overview.

Consider an edit made to join two overlappping audio recordings together. The first recording contains audio up to a point where, for example, a mistake was made; see Figure 1a. The second contains audio starting at a point before the mistake, continuing on to the end; see Figure 1b. To make the edit, the user marks where he wants to go out of the first recording (the "out point") and where he wants to go into the second (the "in point"), see Figure 1c. The edit is performed by playing material from the first take up to the out point and then material from the second take starting at the in point.

In practice automated edit adjustment may be carried out in accordance with this invention as follows.

The user chooses, say, the out point that he wants. He then roughly positions the in point. The audio samples around both in point and the out point are then analysed by calculating a correlation function between the two signals. This should indicate where the best match between the two audio signals occurs and hence the optimum position for the in point. The required adjustment is then either made automatically or indicated to the user.

The automated adjustment can be carried out as follows. A section from each signal (see Figure 3) is divided into blocks of samples. The power spectrum of each of the blocks of samples is then calculated. This produces a series of spectra of the signals at regular time intervals (see Figure 4).

By selecting the same frequency band from each of the spectra, the variation in the power in that frequency band as a function of time is determined (see Figure 5).

The correlation function of the temporal variation of the power in a frequency band from one signal with that from the other has a peak. The position of the peak is related to the temporal shift which, when applied to one signal, brings it into time alignment with the other (for the frequency band in question); see Figure 6.

The position of the peaks of correlation functions from all the frequency bands are collected together (see Figure 7). The best shift to apply to the audio signals to bring them into time alignment is deduced from this assortment of peak positions.

First Embodiment.

Figure 2 shows a suitable implementation of the invention including a disc store 10 holding the two audio recordings to be aligned and an editing facility 11 of known type connected to write to and read from the store. The editing facility 11 makes available the two signals A and B to be compared and supplies them to two fast Fourier transform (FFT) circuits 12 and 13 respectively. Such circuits are commercially available and execute a Fourier transform (or frequency analysis) on the input signal applied thereto. A correlator 14 then compares the outputs of the two FFT circuits, in a manner described below. The output of the correlator is applied to a peak detector 15 which in conjunction with a peak position analayser 16 determines where the peak lies and hence the amount of temporal adjustment required to align the two recordings.

The system of Figure 2 operates as follows. The editing facility outputs two sections of audio data, one from each of the two signals A and B. Typical signal sections are shown in Figure 3. Typically the sections may be 32k (32768) samples long, sampled at a sampling rate of 48kHz, corresponding to two-thirds of a second in duration. The sampling will typically be to 16-bit accuracy. Each 32k sample is then divided in time into 128 blocks each of 256 samples.

The FFT circuits 12, 13 then perform fast Fourier transforms on each of the 128 blocks of each of the two signals to provide for each block a frequency spectrum. Each frequency spectrum will be defined by a block of 128 samples. Figure 4 is a three-dimensional diagram illustrating the two frequency spectra for two typical signals. For each time period corresponding to the duration of one block of the signal the diagram provides a plot of power against frequency. Most of the power is at relatively low frequencies though for signal 1 (as it is here labelled) there is a

notable power component at a relatively high frequency. Figure 4 thus represents the inputs to the correlator 14.

The correlator 14 calculates the correlation function of the temporal variation in each frequency band of one signal with the corresponding variation derived from the other signal. Each spectrum is stored as a 128 word block, and is of the form shown in Figure 5. The power in the first spectral component of each block thus provides a measure of the temporal variation in power for that spectral component, and similarly for the other subsequent spectral frequency components. The correlation function of these two temporal variations in power content of the first spectral band is calculated to find out where variations in power are most alike in the two signals. The correlation function produced is of the type shown in Figure 6. This correlation is carried out by further FFT circuits within the correlator 14. Such correlation is carried out in time for all the spectral frequency components.

The correlation function is:

$$\text{F.T. } \{[F(w)][G^*(w)]\}$$

where F.T. denotes the Fourier transform, the asterisk * denotes the complex conjugate, and F(w) and G(w) are the Fourier transforms of the two time series, i.e. the outputs of the circuits 12 and 13.

During the correlation process it can be beneficial to apply weighting to the functions being correlated. This may be done while the data is in the frequency domain, i.e. after the Fourier transforms of the two functions have been calculated and one has been multiplied by the conjugate of the other, but before the inverse Fourier transform is performed.

An example of such a weighting function is the magnitude squared coherence spectrum which can be considered to be a measure of how much the spectral components of one function are consistent with those of the other function. The spectra of the functions would be divided into segment pairs and the magnitude squared coherence spectrum calculated as follows:

magnitude squared - coherence spectrum (mscs(w))

$$\frac{\left|\sum_{l=1}^{n} F_l(k)\ G_l^*(k)\right|^2}{\sum_{l=1}^{n}\left|F_l(k)\right|^2 \cdot \sum_{l=1}^{n}\left|G_l(k)\right|^2}$$

where spectra F(w) and G(w) have been divided into n segment pairs:-

F0, F1, F2 ..... Fn,
G0, G1, G2 ..... Gn.

Less relevant components of the functions being correlated can be subdued by multiplying the frequency domain data by the magnitude squared coherence spectrum.

The correlation function would be modified:-

F.T. ([F(w)] [G*(w)] [mscs (w)])

The position of the peak in the correlation function of the two arrays shows by how much one array should be temporally offset relative to the other so that they fit best.

In this way 128 plots of correlation against displacement are obtained, one for each frequency band. The peak detector 15 detects the peak of each of these plots. Thus 128 such peak values are obtained, and these are "plotted" as a histogram in the peak position analyser 16 showing how many times a peak occurs at each displacement, as shown in Figure 7. This analyser thus determines the most "popular" displacement of the 128 values obtained for the different frequency bands; this value being used as the required displacement value.

Preferably, rather than just increasing the value in the histogram by one if a peak is found, the size of the peak is added. The size of the peak depends on the original signal amplitude.

Additionally, weighting the different spectral bands may be available as an option to the user, and the range of frequency bands may also be definable by the user. In any event the peak in the resultant histogram is used as the shift required to bring the signals into time alignment for all the frequency bands considered.

The standard deviation of the peak positions plotted on the histogram may be calculated to act as a confidence indicator.

Second Embodiment.

In practice it is convenient to implement the method in a computer or microprocessor as shown in Figure 8 where the special purpose hardware of Figure 2 is replaced by a signal processor 20. The processor 20, which may be a Motorola DSP 56000, operates in accordance with a program which is summarised in the flow chart of Figure 9 which will essentially be self explanatory in view of the description of the first embodiment.

It is particularly convenient to undertake the fast Fourier transforms, required first to produce the frequency spectrum and then in the correlation operation, in the following way. In this method two FFTs can be calculated at the same time when the two signals requiring transforming are both entirely real. One signal is put into the real part of the elements of an array of complex numbers, the other into the imaginary part. When the FFT is performed in place on this array it produces an array containing the complex spectra of both of the signals. The two separate, complex, spectra can be extracted from the array, since the two original signals were entirely real.

Figures 10(a) and (b) show examples of two "real" time series and their Fourier transform (real and imaginary parts).

Figure 10(c) shows the effect of interchanging the real and imaginary parts of a "real" signal.

Figure 10(d) shows the effect of adding together one "real" signal as it is to the other "real" signal with its real and imaginary components interchanged.

A priori knowledge of the even-ness of real part of the Fourier transform of a purely real signal and the odd-ness of a purely imaginary signal makes it possible to extract the real and imaginary

parts of the Fourier transforms of the two original "real" signals as follows:

$$\mathrm{Re}[F(u)] = \mathrm{Re}[\text{F.T. } \{f(t) + i.g(t)\}] + \mathrm{Re}[\text{F.T. } \{f(-t) + i.g(-t)\}]$$

$$\mathrm{Im}[F(u)] = \mathrm{Im}[\text{F.T. } \{f(t) + i.g(t)\}] - \mathrm{Im}[\text{F.T. } \{f(-t) + i.g(-t)\}]$$

$$\mathrm{Re}[G(u)] = \mathrm{Im}[\text{F.T. } \{f(t) + i.g(t)\}] + \mathrm{Im}[\text{F.T. } \{f(-t) + i.g(-t)\}]$$

$$\mathrm{Im}[G(u)] = -(\mathrm{Re}[\text{F.T. } \{f(t) + i.g(t)\}] - \mathrm{Re}[\text{F.T. } \{f(-t) + i.g(-t)\}])$$

where $F(u)$ and $G(u)$ denote the Fourier transform of $f(t)$ and $g(t)$ respectively, and $\mathrm{Re}[x]$ and $\mathrm{Im}[x]$ denote the real and imaginary part of a complex number x respectively.

The methods described, which involve splitting the signals into frequency bands and determining their similarities in the separate frequency bands, have been found to be particularly successful in reliably identifying the amount of displacement required to bring the two signals into alignment.

CLAIMS

1. A method for aligning two audio signals in time, comprising the steps of:

determining the similarity of the two signals for varying time offsets between them, and

deriving from the similarity measurements an optimum time offset to bring the signals into time alignment.

2. A method according to claim 1, in which the similarity of the audio signals is measured in a multiplicity of frequency bands, and the multiplicity of similarity measurements is processed to provide a preferred time offset.

3. A method according to claims 1 or 2, in which when making similarity measurements between any two signals the coherence between those two signals is used to weight the similarity measurements of the signals.

4. A method according to any preceding claim, in which the power of the harmonics in the various frequency bands is used to weight the similarity measurements.

5. A method according to any of claims 2 to 4, in which the similarity measurements are weighted according to frequency band.

6. A method according to claim 1, in which the signals are divided into blocks, the frequency spectrum of each block is determined, the power variation with time is determined for each of a plurality of frequency bands, the power variation of the two signals is correlated for each frequency band, the peak of each correlation function is determined, and the peak value of the peaks thus obtained determined to provide a desired offset.

7. A method of aligning two audio signals in time, substantially as herein described with reference to the drawings.

8. Apparatus for aligning two audio signals in time, comprising:

means for determining the similarity of the two signals for varying time offsets between them, and

means for deriving from the similarity measurements an optimum time offset.

9. Apparatus according to claim 8, in which the similarity of the audio signals is measured in a multiplicity of frequency bands, and the multiplicity of similarity measurements is processed to provide a preferred time offset.

10. Apparatus according to claim 9, in which the power of the components in the various frequency bands is used to weight the similarity measurements.

11. Apparatus according to claim 9 or 10, in which the similarity measurements are weighted according to frequency band.

12. Apparatus according to claim 8, in which the signals are divided into blocks, and including means for determining the frequency spectrum of each block, means for determining the power variation with time for each of a plurality of frequency bands, means for correlating the power variation of the two signals for each frequency band, means for determining the peak of each correlation function, and means for determining the peak value of the peaks thus obtained to provide a desired offset.

13. Apparatus for aligning two audio signals in time, substantially as herein described with reference to the drawings.

14. A method of editing audio signals, including aligning the audio signals by a method in accordance with any of claims 1 to 7.

15. Audio signal editing apparatus, including apparatus for aligning two audio signals in accordance with any of claims 8 to 13.

Published 1990 at The Patent Office, State House, 66/71 High Holborn, London WC1R 4TP. Further copies may be obtained from The Patent Office. Sales Branch, St Mary Cray, Orpington, Kent BR5 3RD. Printed by Multiplex techniques ltd, St Mary Cray, Kent. Con. 1/87